AD_____

AWARD NUMBER: DAMD17-02-1-0350

TITLE: "Identification of the Downstream Promoter Targets of Smad Tumor Suppressors in Human Breast Cancer Cells"

PRINCIPAL INVESTIGATOR: Xuedong Liu, Ph.D.

CONTRACTING ORGANIZATION: University of Colorado
Boulder, Colorado 80291-0220

REPORT DATE: July 2006

TYPE OF REPORT: Final Addendum

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-07-2006	Final Addendum	1 Jul 2005 – 30 Jun 2006

4. TITLE AND SUBTITLE

"Identification of the Downstream Promoter Targets of Smad Tumor Suppressors in Human Breast Cancer Cells"

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-02-1-0350

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Xuedong Liu, Ph.D.

E-Mail: liux@colorado.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Colorado
Boulder, Colorado 80291-0220

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Members of TGF-ß superfamily ligands are potent regulators of cell proliferation, differentiation and animal development. The biological effects of these ligands are mediated mainly through activation of the Smad family protein which serves as transcription factors to regulate gene expression. We analyzed the cis-regulatory elements that are responsible for conveying TGF-ß/Activin responses at the genomic levels. Activin A and TGF-ß transcriptional responses in immortalized normal human mammary epithelial cells were examined by gene expression profiling and computational analysis the regulatory regions of TGF-ß-responsive genes using a new algorithm, which is based on frequency of occurrence, and cross-species conservation. Our analysis revealed that a distinct set of cis-regulatory elements conserved across species is either unique or over-represented in TGF-ß-regulated genes. A set of bioinformatics tools were developed for browsing promoters in mammalian cells.

15. SUBJECT TERMS
No subject terms provided.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	13	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

Progress report:

1. INTRODUCTION

Members of transforming growth factor ß (TGF-ß) superfamily play important role normal mammary gland development and serves as tumor suppressor function. Alteration of transforming growth factor ß (TGF-ß) signal transduction pathway is one of the key cellular events in the pathogenesis of breast cancer. The roles of TGF-ß family members play during normal cell proliferation and differentiation have not been fully characterized. Components in the TGF-ß signal transduction pathway are frequently mutated in breast cancer cells. For example, mutations in TGF-ß type I receptor are detected in metastatic breast cancers. Smad4, which is one of the intracellular signaling molecules that transduces the TGF-ß signal from the cell surface into the nucleus, is deleted in several breast cancer cell lines. The goal of our investigation is to understand the molecular mechanism of tumor suppression by TGF-ß by identifying the downstream promoter targets of Smads tumor suppressors in normal and breast cancer cells. We have made significant progress in our study and identify genes that are responsive to TGF-ß and its related ligand Activin in normal human mammary epithelial cells. In addition, we have identified genes whose responses are likely to be dependent on the presence of Smad4 tumor suppressor gene. Finally, a set of bioinformatics tools has been developed for genome wide analysis of cis-regulatory codes.

2. BODY---Studies and Results

Three specific aims were proposed in the original application:

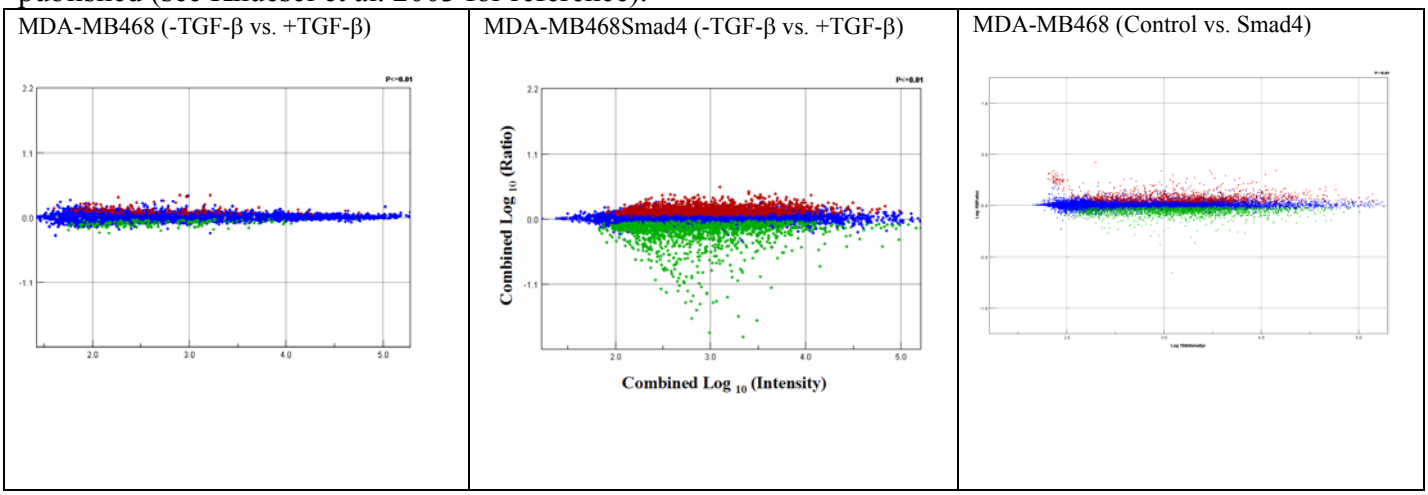
1. *Development of a novel chromatin immunoprecipitation assay (CHIPS) using a TAP-TAG system to isolate in vivo binding targets of Smad3 and Smad4.*
2. *Identification of the downstream promoter targets of Smad3 or Smad4 in breast cancer cells.*
3. *Identify Smad4 regulated downstream target genes in tumor cells using DNA microarray technology*

Task 1. Development of a novel chromatin immunoprecipitation assay (CHIPS) using a TAP-TAG system to isolate in vivo binding targets of Smad3 and Smad4, ***(months 1-24)***

- *grow sufficient quantity of MDA-MB468 cell lines for CHIPS analysis* ***(months 1-2).***
- *Optimize the experimental procedure for two step purification of TAP tagged Smad3 or Smad4 from cell lysates* ***(months 3-5)***
- *Optimize the crosslinking and sonication conditions for Smad3 and Smad4* ***(months 6-8)***

We have constructed and characterized human breast cancer cell lines expressing TAP-Tag Smad3 and Smad4. We have done preliminary DNA microarray analysis on these two cell lines. We compared expression profiles of Smad4 null MDA-MB468 cell line with the same cell line in which we stably expressed Smad4. In addition, expression profiles of each cell line treated with or without TGF-beta were also analyzed. RNAs were extracted from control and MDA-MB468 cell lines stably expressing TAP-tag Smad4. Eight replicates were done for each comparison. Agilent human 1 cDNA microarray for our expression profiling analysis. As shown in Figure 1, the effect on gene expression by the presence or absence of Smad4 expression or treatment with TGF-beta were measured. Data points represent Resolver-combined log ratios for differentially expressed genes. Log ratios colored blue are unchanged, those shown in red are up-regulated and those in green are down-regulated. We can draw two conclusions from this initial study. 1) breast cancer cell line MDA-MB468 exhibits limited response to TGF-beta signaling even in the absence of Smad4. 2) When Smad4 expression is restore, more robust transcription response to TGF-beta is observed. Therefore, there are a number of Smad4-dependent genes. It is possible to identify Smad4-dependent and TGF-beta-dependent gene.

We have optimized the experimental procedure for two-step purification of TAP tagged Smad3 or Smad4 from cell lysates. The detailed procedure for TAP purification of Smad3 protein complex has been published (see Knuesel et al. 2003 for reference).



We optimized the crosslinking and sonication conditions for Smad3 and Smad4 using the PAI-1 promoter region in our CHIP assay. However, we still see a significant contamination of non-specific DNA fragment coming down in our CHIP experiment. The signal to noise ratio has not changed significantly when we trying to vary the condition of sonication, different types of beads for immunoprecipitation. We have also changed our protocol of CHIP in hope to improve the signal to noise ratio. After sonication, we loaded the soluablized chromatin onto a noncontinous CsCl gradient to purify to chromatin fraction away sonicated DNA fragments and RNAs. This purification does not appear to improve signal to noise ratio significantly. We have tried to clone the fragment from CHIP assay and sequenced more than a dozen of these fragments. None of those fragments passed the secondary screen using Smads protein complex for gel-shift assay.

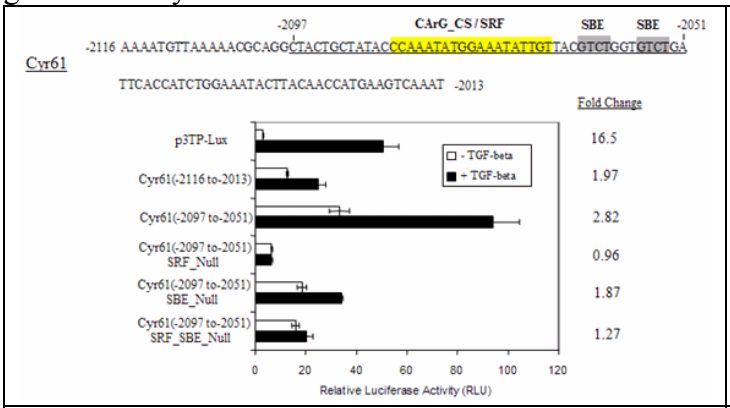


Figure 2A. A schematic representation of the Cyr61 promoter region identified by computational analysis was shown above the graph with known transcription factor binding sites highlighted. Regions of Cyr61 promoter were cloned into luciferase reporter gene and mutagenesis was performed with SRF and SBE sites. Transcriptional responses of the indicates reporters were measured after transfections in Hep3B cells. The fold induction by TGF-ß is indicated and error bars represent standard deviations from triplicate determinations.

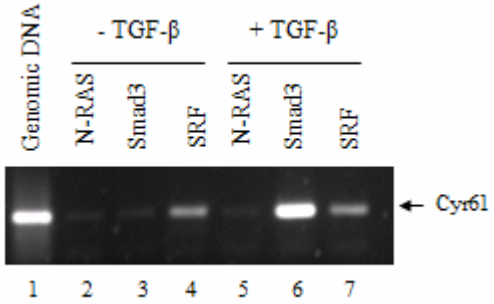


Figure 2B. SRF and Smad3 associate with Cyr61 promoter *in vivo*. Human mammary epithelial cells treated with or without TGF-ß for 2 hr were cross-linked using 1% formaldehyde for 15 min. Chromatin immunoprecipitation experiment was performed as described in Materials and Methods. Genomic DNA isolated from HME cells was used as control for PCR amplification. Anti-N-Ras antibody was used a negative control for nonspecific DNA binding.

Our discouraging initial attempt to identify the bona fide Smad3 or Smad4 binding sites by CHIP assay following by direct cloning approach forced us to rethink about the most efficient way to accomplish our goal---identification of the downstream promoter targets of Smad tumor suppressor. Since CHIP assay can successfully recover the binding site in the promoter regions of Smad-dependent TGF-beta regulated genes in a mixture of IPed fragment, the success rate would be higher if we know Smad-dependent TGF-beta regulated genes. Based on the knowledge of Smad3 and Smad4 binding behaviors that documented in a number of studies, it is possible to analyze the promoter regions of those genes by bioinformatics and

subsequently confirm the Smad association in the promoter region by CHIP assay. Indeed, we have demonstrated that we can effectively predict novel TGF-beta responsive elements using the bioinformatics tools we developed and validate that Smads bind to these elements using CHIP assay. Shown in Figure 2 is a novel composite TGF-beta responsive element containing SRF and SBE. Smad3 and Smad4 clearly bind to this region upon TGF-beta stimulation.

Task 2. Identification of the downstream promoter targets of Smad3 or Smad4 in breast cancer cells ***(months 20-48)***

1) *Workout ligation mediated PCR protocol for amplification of unknown targets of Smad3/4 binding sites* ***(months 20-24)***
2) *Cloning of the amplified Smad3/4 binding sites into a luciferase reporter construct* ***(months 25-28)***
3) *Make small pool library of the cloned putative Smad3/4 binding sites. Pool size=10. Initial plan is to make 100 pools* ***(months 29-32)***
4) *Transient transfection of HepG2 cells each small pool and screen for TGF-ß responsive pools* ***(months 33-36)***
5) *Subdivide each positive pool to identify individual clone that mediate TGF-ß transcriptional response* ***(months 36-38)***
6) *Sequence each positive clone and obtain the identity of the genes that are regulated by TGF-ß through the binding site* ***(months 39-42)***
7) *Confirm the binding of the identified DNA fragment to purified Smad3 or Smad4 in vitro by a gel shift assay* ***(months 43-45)***
8) *Mutational analysis to confirm the importance of the Smad binding site in mediating TGF-ß transcriptional response* ***(months 43-48)***

We have performed the subtask 1-4 in Task 2. Unfortunately, we did not obtain the intended results as we had hoped. We decided to pursue an alternative strategy to achieve the goal of Task 2, i.e. identification of the downstream promoter targets of Smad3 or Smad4 in breast cancer cells. We hypothesized that a limited set of cis-regulatory elements alone or in combination in conducting TGF-ß transcriptional responses. Some of these regulatory elements bind Smads directly. It has been demonstrated that TGF-beta induced transcriptional responses are conserved among human, mouse and rat. We would expect that cis-regulatory elements of TGF-ß responsive genes would be conserved across these species. Computational analysis of promoter regions of TGF-beta responsive genes would yield a list of candidate transcription factor binding sites that are likely to mediate TGF-beta transcriptional responses. Therefore, we decided to determine Activin A and TGF-ß transcriptional responses in immortalized normal human mammary epithelial cells by gene expression profiling. We performed computational analysis the regulatory regions of TGF-ß-responsive genes using a new algorithm we developed, which is based on frequency of occurrence, and cross-species conservation.

Development of GeneACT

The overview of GeneACT is summarized in Figure 3. Genomic sequence data from human, mouse, and rat, Transcription Factor Database (TFD) and orthologs information (NCBI HomoloGene) are downloaded from NCBI. GeneACT is built on top of a PostgreSQL database. To facilitate the differential binding site search (described below), we stored the occupancies of all the binding sites in the TFD database (approximately 7000 known binding sites) in each gene found in a HomoloGene group that spans all three species up to 10000 bp upstream from the start codon. Users can access the database via the GeneACT web interface

at http://promoter.colorado.edu/geneact. For the most in-depth information on how to use GeneACT, help documentation is available (http://promoter.colorado.edu/geneact/help.html).

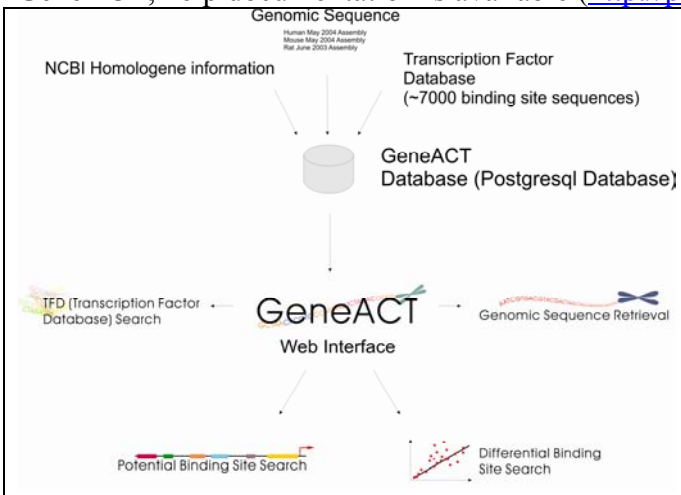


Figure 3 Overview of the GeneACT architecture and method

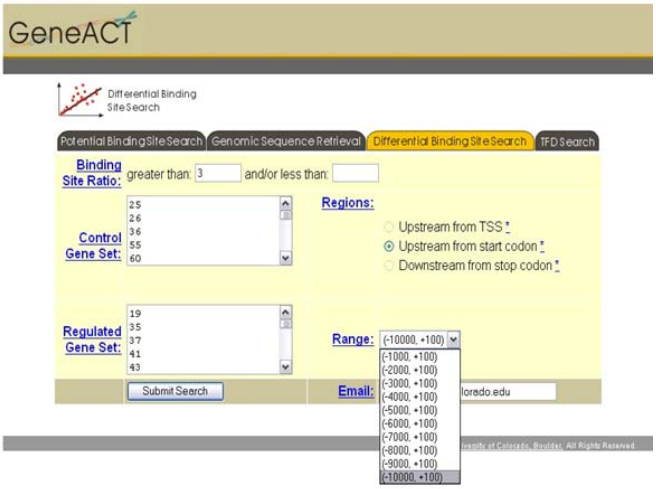


Figure 4. Web interface of Differential Binding Site Search. Gene IDs from control gene set (unchanged in DNA microarray data) and regulated gene set (up- or down- regulated from microarray data) are pasted into respective windows. The threshold of binding site ratio is defined by the user. The user can specify a range of interest with three choices of regions (upstream from the transcription start site, upstream from the start codon, downstream from the stop codon).

Using Gene Expression Data to Discover Networks of Transcription Factors Using Differential Binding Site Search (DBSS)

The use of microarrays to elucidate genomic scale gene expression patterns is now widespread. These microarray experiments generate large sets of differentially expressed genes, yet the actual mechanism that controls the differential gene expression cannot be readily deduced using this technique. It is well known that specific transcription factor binding elements in the promoter region are largely responsible for differential gene expression. However, the short, degenerate nature of these binding sites leads to an unacceptably high false positive rate during computational searches in the promoter regions. Furthermore, these binding sites often only respond under specific treatments or conditions, making it extremely difficult to predict the biological significance of computationally determined binding sites within the promoter. Because of the differential patterns observed in the microarray studies, it is our expectation that there is a differential distribution of regulatory sequence elements between the differentially expressed genes compared to that of the control genes in a particular system. To gain global insight into the mechanism involved in a particular system in regards to what transcription factors are involved, Differential Binding Site Search (DBSS) in GeneACT was created (Figure 4).

DBSS takes as input two sets of genes: a control set and a regulated set. It then calculates frequencies of genes that contain such binding sites found in the regulated set and control set genes and the fold change in frequency of each binding site. DBSS has been designed to discover transcription factor binding sites that are enriched in the regulated gene set. The control gene set is used to determine a baseline for background noise. Each binding site that is found in a regulatory region that spans the human, mouse, and rat genomes is counted in the final output. At present, we preprocessed the -10000 bp to +100 bp region of each gene that contains ortholog information in NCBI HomoloGene centered across the start codon of each gene. Although the binding site sequences in the TFD are all experimentally determined in the literature, many of these sequences are short, , and overlapping. Despite the fact that restricting the binding sites to just those that span multiple genomes greatly reduces the overall background noise, certain short degenerate binding site sequences may

still appear as false positives. Thus, the use of the control set of genes is crucial to further reduce the false positive rate. For binding sites that do not contribute to the regulation of a particular gene, we expect there to be no relative change in frequency. These genes are then filtered from the results by specifying a lower bound for the "Binding Site Ratio" option on the search interface. For example, to keep only the binding sites that have three times the frequency in the regulated set versus the control set, you would specify a lower bound of three. By looking at the binding sites that have a large ratio (fold change) between the regulated set genes and control set genes, the binding site sequences that are potentially important to the regulation of a given system under specific conditions or treatments can quickly be determined. In this way, the regulatory mechanism of how the transcription factors regulate a given system can be inferred from the enriched binding site sequences.

Transcriptomic profiling of TGF-ß and Activin A responses in HME cells

To uncover genes that are regulated by TGF-beta signaling, we performed DNA microarray experiments using human mammary epithelial cells (HME). To gain further insight into TGF-ß and Activin A signaling response in the relatively normal mammary epithelial cells, we compared TGF-ß- and Activin A-regulated time-dependent gene expression patterns in HME cells. Total RNA was isolated from human mammary epithelial cells treated with 100 pM TGF-ß or Activin A for 2, 4 and 8 hours. To assess the changes in relative abundance of transcripts in response to TGF-ß and Activin A treatment, total RNA from non-treated control cells (T0 for control cells that were not treated with TGF-ß and A0 for Activin A non-treated cells) or treated cells were amplified and labeled with either Cy3 or Cy5 fluorescent dyes. In each experiment, Cy3-labeled amplified RNA (aRNA) from non-treated cells was mixed with Cy5 labeled amplified RNA derived from TGF-ß or Activin A treated cells at the indicated time points and hybridized to Human 1A 60-mer oligonucleotide arrays representing more than 17,000 human genes. Each experiment consisted of four replicates of hybridization.

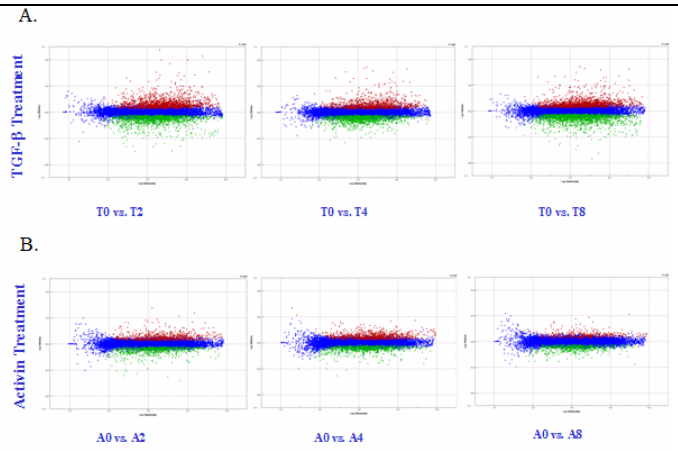


Figure 5. Comparison of time-dependent expression profiles on human mammary epithelial cells (HME) treated with either TGF-ß (A) or Activin A (B) for indicated times. All data points represent combined values across four replicate arrays. Log ratios colored blue are unchanged (not significantly different than 0), those shown in red are up-regulated (significantly greater than 0, p value <0.01) and those in green are down-regulated (significantly less than 0, p value <0.01). The data were processed by the Resolver software (Rosetta Biosoft) and plotted - Log(10) Ratio vs. Log(10) Intensity.

	Activin A		TGF-beta		Common	
	Up	Down	Up	Down	Up	Down
2 hour	22	61	126	114	12	10
4 hour	24	40	87	69	6	20
8 hour	14	24	92	151	5	13

Figure 6. Summary of the changes of gene expression profiles determined by DNA microarray array analysis.

Computational Analysis of cis-regulatory Sequence Elements in the Promoter Regions of TGF-ß Responsive Genes

Our transcriptomic profiling of HME cells treated with TGF-ß/Activin A led to the identification of groups of ligand-responsive genes. What are the cis-regulatory elements embedded in the control regions of these genes that are most likely responsible for TGF-ß induction or repression? Whether a number of TGF-ß-responsive genes share similar cis-regulatory elements are still largely unknown. We took a comparative genomic approach and use GeneACT to define cis-regulatory elements that are unique or overrepresented in the promoter regions of TGF-ß-responsive genes. We used a set of 108 genes that are differentially expressed upon TGF-beta stimulation (at least 1.8 fold induction or repression at the 2 hour time point) and a set of genes that are not regulated by TGF-beta (fold changes on microarray in between -0.001 fold to 0.001 fold in all four replicates) to search for all binding sites of these genes in their promoter regions upstream from the transcription start site (TSS) and translation start codon. We hypothesized that the frequency of the TGF-ß-responsive binding sites present in the TGF-ß-regulated genes is significantly higher or lower than that of the unregulated genes. To examine this we used a set of 644 unregulated genes as our control set to reflect a basal frequency of a particular binding site occurrence in the genome upon ligand treatment. 108 TGF-ß regulated genes were also chosen and the frequency of each of the transcription factor binding sites existing in the TFD was calculated. In addition, those transcription factor binding sites that occur more frequently in the regulated gene set than in the control set (>= 2.9 fold) are also documented.

Experimental Validation and Characterization of Two Novel TGF-ß-Responsive Elements Predicted from Computational Analysis

Our computational analysis suggested a collection of potential TGF-ß- responsive elements in the genome. Whether any of these elements other than the ones that are well-characterized in the literature make biological sense remains to be determined experimentally. Since the SRF binding site is highly overrepresented and shared by many TGF-ß-responsive, we first sought to validate SRF as a bona fide TGF-ß-responsive element. To begin with, we chose two TGF-ß target genes, CYR61 and HS3ST2 from our microarray list. The regulatory elements that are responsible for TGF-ß-responsiveness in the promoter regions of these two genes have never been characterized. Data presented in Figure 2A implicated the region between -2116 and -2013 in CYR61 and -1070 and -1030 in HS3ST2 to be involved in mediating TGF-ß responses. Another reason for selecting these two regions is that the nucleotide sequences of these regions are conserved between human, mouse and rat. The indicated regions (Figure 2A) were cloned into a luciferase reporter construct (pGL3). To test whether the promoter fragment containing -2116 to -2013 of CYR61 can confer a TGF-ß response, the reporter construct was transfected into Hep3B cells. Hep3B cells are highly transfectable and have been used as model cell lines for analyzing TGF-ß signaling. HME cells are less transfectable and TGF-ß transcriptional responses in HME are transient, thus making it difficult to perform reporter gene assays. As positive controls, p3TP-Lux and p3APP-Lux, two standard TGF-ß signaling reporters, were also transfected. As shown in Figure 2A, the region spanning -2116 to -2013 is able to confer a modest TGF-ß response in Hep3B cells (1.97-fold increase). A consensus SRF binding site is located between -2085 and -2067. To test whether this SRF site is responsible for TGF-ß induction, a pair of oligos containing the SRF sequence (underlined) was inserted into pGL3 (Figure 2A). In the presence of TGF-ß, this reporter gene showed 2.82-fold activation indicating the SRF sequence element could be responsible for mediating TGF-ß induction. Further inspection of the sequence in this region revealed that two potential Smad binding elements (SBE) in tandem are located downstream of the putative SRF element. To assess the relevance of these sequence elements in TGF-ß gene induction, mutations that have been

previously shown to prevent SRF and Smad binding to these sequence elements were introduced in this region individually or in combination. Mutation of the putative SRF binding element eliminates TGF-ß induction and causes a modest reduction in the basal level of transcription, in contrast, mutation of the two SBE elements reduces but not completely impairs TGF-ß-responsiveness. Finally, combined mutations of the SRF and SBE result in negating much of TGF-ß-induced gene activation. These results indicate that the SRF element but not the SBE is the primary TGF-ß-responsive element. SBE can enhance TGF-ß-responsiveness only in conjunction with SRF. Without SRF, the tandem SBE is unable to mediate transcriptional response to TGF-ß.

To further prove that the putative SRF and SBE elements in the Cyr61 promoter region binds the SRF and Smad3 transcription factors in vivo, we performed chromatin immunoprecipitation analysis (ChIP) on HME cells using antibodies against SRF and Smad3. An antibody raised against N-Ras was used as a control. As shown in Figure 2B, SRF was found to bind the region -2116 to -2013 of Cyr61 regardless of ligand stimulation (Figure 2B). However, Smad3 only binds this region upon treatment with TGF-ß. Thus, SRF and Smad3 are likely to be transcription factors that regulate gene induction of Cyr61 promoter.

The gene encoding heparan sulfate (glucosamine) 3-O-sulfotransferase 2 (HS3ST2) was found to be a TGF-ß-regulated gene in this study. The region spanning -1070 and -1030 was found to be the one containing the candidate regulatory elements by the computational analysis. Within this region there is a TRE element (GTGAGTCAG) and a potential Smad binding element (SBE) (Figure 7). To test effectiveness of this relatively small region to enable TGF-ß induction, reporter constructs consisting of one, two or three copies of this elements were made and transfected into Hep3B and mink lung cells. The results shown in Figure 7 are data obtained with Hep3B; similar results were obtained with mink lung cells. A single copy of this element is able to elicit a 2.84-fold of activation in the presence of TGF-ß. As the copy number of this responsive element increases, so does TGF-ß induction as well as the basal levels of transcription. This result indicates that this 40 bp sequence consisting of the TRE element next to AGAC is a TGF-ß-responsive element for HS3ST2. Taken together, our experimental studies in the two cases we investigated support our computational predictions. Further experiments will be necessary to validate other candidate elements in an effort to fully categorize the cis-acting regulatory elements responsible for TGF-ß induction.

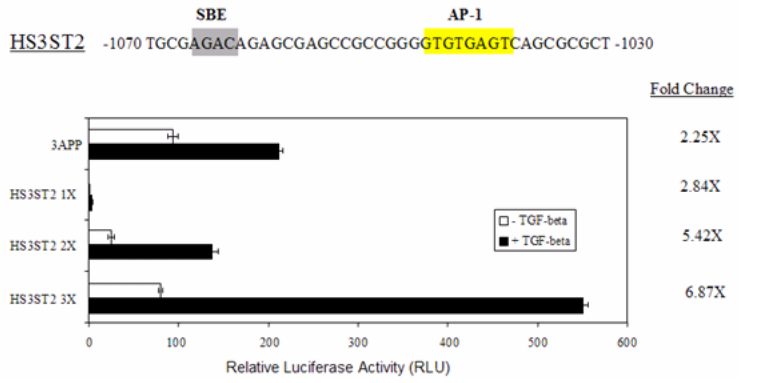


Figure 7. Hep3B cells were transfected with reporter constructs containing one, two or three copies of the putative minimal TGF-ß responsive element from the promoter region of the HS3ST2 gene. p3APP-Lux was used as the positive control in this experiment.

In summary, we have developed an alternative approach to identify TGF-beta/Smad binding sites in TGF-beta regulated genes. We have accomplished the task 2 through a more effective approach.

Task 3. Identify Smad4 regulated downstream target genes in tumor cells by DNA microarray ***(months 12-30)***

1) *prepare high quality of mRNA for DNA microarray analysis* ***(months 12-14)***
2) *run test chip experiment to familiar with the procedure and calibrate the reagent* ***(months 15-16)***
3) *Prepare high quality cRNA for hybridization to the U95 CHIP* ***(months 17-18)***
4) *Hybridization, scan and data collection* ***(months 19-20)***
   Analysis the DNA microarray data using gene spring or cluster software ***(months 20-24)***
5) *Annual report will be written* ***(months 20-24)***
6) *Repeat the DNA microarray experiment to ensure the high reproducibility of the data* ***(months 25-30)***
7) *Summary of DNA microarray data will written and initial manuscript will be drafted* ***(months 25-30)***

We have constructed several pairs of cell lines that differed in the expression of Smad4. These include MDA-MB468 breast carcinoma, SW480 and CFPAC-1 cells. We have compared the expression patterns between these pairs of cell lines in the presence and absence of TGF-ß using a DNA microarray technique. This analysis has revealed that there are Smad4-dependent and Smad4-independent TGF-ß regulated genes. However, we were very surprised to find that there is little overlap among these cell lines i.e. Smad4-dependent and Smad4-independent TGF-ß regulated genes are different for each pair of cell lines. These results suggest that Smad4-dependent and Smad4-independent genes are highly cell type and context-dependent. This reinforced our notion that Smads are unlikely the solely important transcription factors in determining the outcome the transcriptional response. TGF-beta induced gene expression is highly cell type and context-dependent suggesting that TGF-beta inducible gene expression is highly combinatorial in nature.

3. KEY RESEARCH ACCOMPLISHMENTS

1. Obtained gene expression profiling data in human mammary epithelial cells in response to TGF-beta and Activin A.
2. Obtained gene expression profiling data in human mammary epithelial cells in response to various concentrations of Activin A.
3. Construct a human, mouse and rat promoter database for bioinformatics analysis of TGF-ß responsive promoters.
4. Obtained a complete dataset for the regulatory elements in the promoter regions of the TGF-beta responsive genes that conserved across human, mouse and rat genome.
5. Identified two novel TGF-beta responsive elements that are responsible for TGF-beta induced transcriptional activation of Cyr61 and HS3ST2
6. Obtained gene expression profiling data in MDA-MB468, SW480 and CFPAC-1 and their Smad4 expressing derivative tumor cell lines in response to TGF-beta.
7. Developed a Web-based Promoter Browser for Gene Expression Analysis
8. Developed a Method for Employing Gene Expression Data to Discover Networks of Transcription Factors Using Differential Binding Site Search (DBSS)

4. REPORTABLE OUTCOMES

Manuscript

Cheung, H.T., Collins, P.J., Gao, Y., Riquelme, C., Kwan, P., Doan, T.B and X.Liu Specificity of TGF-beta and Activin Signaling Responses Revealed by the Analysis of Their Transcriptional Programs. Submitted to Molecular Cellular Biology and *in revision*.

Cheung, T., Y. Kwan, M. Hamady, and X. Liu. Unraveling transcriptional control and *cis*-regulatory codes using GeneACT. *Genome Biology and in revision. Provisionally accepted.*

Cheung, H.T., Collins, P.J., Kwan, P., Doan, T.B and X.Liu Comparison of TGF-beta and Activin A signaling specificity in breast and liver cell lines. *Manuscript in preparation.*

Bioinformatics Tools and Database

A Web-base Bioinformatics Tool Package for Gene Expression Analysis.
(http://promoter.colorado.edu)

Presentation

July, 2006	Invited conference speaker, The First International Conference on Computational Systems Biology, Shanghai, China, 2006.
Oct. 2005	Invited conference speaker, Biotechnology and Bioinformatics Symposium, Colorado Springs, CO
July 2005	Invited conference speaker, Society for Developmental Biology (SDB) 64th Annual meeting in San Francisco, CA
June 2005	Invited seminar speaker, University of Pennsylvania Ambrason Cancer Center. Philadelphia, PA
June 2005	Conference speaker, FASEB Summer Research Conference on TGF-beta signaling. Snowmass, CO
June 2005	Poster presentation, Era of Hope, BRCP meeting, Philadelphia, PA
Mar. 2005	Invited seminar speaker, University of Colorado Cancer Center, Denver, CO
July 2004	Poster presentation, Keystone Symposium on Bioinformatics, Streamboat Spring, CO

Training and Degree

Tom Cheung: Ph.D. in Chemistry, May 2006. Currently Postdoctoral Fellow in Tom Rando's laboratory at Stanford University.

Phoenix Yin Kwan, M.S. in Computer Science and minor in Biochemistry, Dec. 2005. Currently employed at Dharmacon Inc.

5. CONCLUSIONS

The objective of this proposal is to identify the downstream transcription targets of Smad tumor suppressors in breast cancer cells and characterize heritable changes in tumor cells due to the deletion of Smad4 using an innovative technique. We have achieved our goal and identify TGF-beta response genes in normal breast and breast cancers cells. Along the way, we have developed novel and innovative bioinformatics tools and technologies that will have broad applicability in studying gene expression in mammalian cells. A web-based cis-acting element

browser (GeneACT) which provides graphic visualization and extraction of common regulatory codes in the promoters and 3'-UTRs that are evolutionarily conserved across multiple mammalian species in a particular biological context is described. Using the tools we developed, we analyzed TGF-beta induced transcriptional responses in normal and cancer breast cells using DNA microarray. We have identified transcription factor binding sites that are likely to be involved in mediating TGF-beta transcriptional response. Furthermore, we have validated experimentally two novel TGF-beta responsive elements and demonstrated that Smads bind these elements *in vivo*. We have submitted our results for publication and are now in the process revising our manuscript to address the concerns of the reviewers in hope to get them publish soon.